Supporting Information

Weichselbaum *et al.* 10.1073/pnas.0809242105

SI Text

Study Populations. Previously, we reported on gene expression profiles of tumors from a series of 295 patients with stage I/II breast cancer (NKI295) treated at the Netherlands Cancer Institute between 1984 and 1995 (1). The clinical data used for the earlier publications was updated until January 2001. For this study, all patient charts were reviewed and clinical data were updated until January 1, 2005. The median follow-up times are 10.2 years for all patients and 12 years for patients who were alive. Distant metastasis was analyzed as first event only. If a patient developed a local recurrence, axillary recurrence, contralateral breast cancer, or a second primary cancer (except for non-melanoma skin cancer), the patient's data were censored at this time. An ipsilateral supraclavicular recurrence shortly preceded distant metastasis in all but one patient; therefore, these cases were not censored at time of ipsilateral supraclavicular recurrence. There were 161 patients who underwent breast conservation that consisted of adjuvant external-beam radiation primarily to 50 Gy (mean, 50.2 Gy; range, 50–54 Gy) followed by a boost (89% of patients) using photons, electrons, or iridium Ir 192 (mean, 18 Gy; range, 14–26 Gy). There were 110 patients who received adjuvant chemotherapy, which primarily consisted of cyclophosphamide/methotrexate/5-fluorouracil. Two patients were treated with 5-fluorouracil/epirubicin/cyclophosphamide. Table S3 lists the patient characteristics.

Details on the clinical characteristics of other patients and microarray data used in the study along with accession numbers have been described. References for all microarray data used in this study are listed in Table S8. Available patient characteristics for the patient data sets used in the validation studies are given in Tables S5 and S7. For more information on the patients used in the validation studies, see *Validation of the IRDS*.

Cell Line and Animal Experiments. For *in vivo* experiments, nude mice were treated by i.p. injection with 6 mg/kg of doxorubicin. For lung metastasis assays, gross lung lesions were scored after necropsy at 9 to 12 weeks. *In vitro* growth inhibition was measured by MTS assay and cell death was measured by subG1 content using flow cytometry and confirmed by PARP cleavage assay (Invitrogen). For stable gene knockdowns, Nu61 was retrovirally transduced with either the pSIREN-RetroQ-DsRed-Express (Clontech) to target STAT1 by shRNA (9). The pSHAG-Magic v2.0 vector (Open Biosystems) was used to independently target STAT1 and to target IFIT1 and ISG15 (see Table S10 for sequences). For STAT1 over-expression, the SKBR3 human breast cancer cell line was transduced with the pQCXIP retroviral vector (Clontech) containing a human STAT1α cDNA.

Determination of Clinical Information, Prognostic Marker Status, and Risk Stratification. Most of the clinical and pathological information for the 295 patients has been previously published (2). Molecular subtype assignments are from Fan *et al.* (3). Estimation of Her2 amplicon expression using the microarray data were done using the probes for Her2/ERBB2 and GRB7. For this, hierarchical clustering using Euclidean distance as the distance metric and complete linkage was used, and the resulting tree was cut to give three groups. Cutting the tree at this level gave the highest R-index of 0.867, which is a measure of robustness (4). Clusters 2 and 3 demonstrated the highest expression levels for Her2/ERBB2 and GRB7 and included 26% of the population, which is the expected approximate frequency for Her2 over-expression by FISH and/or IHC. Survival analysis comparing clusters 2 and 3 versus cluster 1 revealed an expected difference in metastasis-free survival. Therefore, patients in clusters 2 and 3 were used for an estimate for Her2 over-expression. Stratification into clinical risk groups was based on the 2005 St. Gallen consensus criteria (5) and using the microarray data for Her2 over-expression. Data on lymphovascular invasion were not available. For AOL 10-year breast cancer mortality estimates, clinicopathological factors were entered into the Web-based tool, version 8.0. For co-morbidity status, "average for age" was used.

Hierarchical clustering of clinical, pathological, and genomic markers was performed using the Heatplus package 1.2.0 (by Alexander Ploner) for the R language and environment for statistical programming, versions 2.31 to 2.4.1 (R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria). Statistical association between various prognostic markers and risk groups was calculated with a χ^2 test and/or Cramer V statistic using the vcd package 1.0.4 for R (by David Meyer, Achim Zeileis, and Kurt Hornik).

Gene Set Enrichment Analysis of IRDS with SF2. Radiation resistance data for 34 of the NCI60 cancer cell lines has been previously described (6). Affymetrix U133A microarray data processed using the RMA method was downloaded from the NCI/LMP Genomics and Bioinformatics group (http://discover.nci.nih.gov/cellminer/home.do). Gene set enrichment analysis was performed using GSEA 2.0 downloaded from the GSEA Web site (http://www.broad.mit.edu/gsea/index.html). A Pearson correlation to the SF2 was used as the metric. Permutation of samples was used to calculate significance and false discovery rate. Analysis was performed by using all probe sets or by collapsing probe sets to unique gene symbols with similar results. Similar results were obtained with gene set analysis using the GSA 1.0 package for R (by Brad Efron and R. Tibshirani). The max-mean statistic was used and re-standardization was performed using all genes in the data set.

Analysis of IRDS Expression in Primary Human Tumors. The human cell line SCC61 was xenografted into the flank of immunocompromised nude mice and *in vivo* selected for resistance to DNA damage as previously described (7). This selection resulted in the resistant Nu61 subline, which was compared with the SCC61 cell lines for differential microarray gene expression analysis using the Affymetrix U133A GeneChip. From this, a 54-gene signature associated with DNA damage resistance was developed as previously described, and duplicate probes were removed to give a final list of 49 unique genes (Table S9). Using this gene list, the average signal intensity for each gene was computed from triplicate samples from SCC61 and Nu61 and transformed into log base 2.

To relate a cell line-derived gene expression signature to primary human tumor samples, we used similar methods described in our previous work (8). For non-Affymetrix platforms (i.e., breast, head and neck), the corresponding probes for each of the 49 IRDS genes were matched based on gene symbols and UniGene accession numbers and duplicate probes removed (Table S9). Using the IRDS genes, k-means clustering was performed using TIGR MultiExperiment Viewer version 4.0 (9) for each of the microarray data sets with a k value of 2 and requiring 90% consensus for each of the two clusters after 500 runs. The average signal intensity for each of the IRDS genes was

then averaged for each of the two consensus clusters. These centroids were then compared with the Nu61 and SCC61 cell line centroids by using a Pearson product moment correlation coefficient. The patients in the group with the centroid positively correlating to the Nu61 centroid were defined as IRDS(+) and those in the group positively correlating with the SCC61 centroid were defined as IRDS(−). The few patients not assigned to a consensus cluster were considered IRDS(−).

The gene expression profile of the IRDS genes for each patient sample from the various primary tumor types was also directly compared with the IRDS gene profiles of Nu61 samples ($n = 6$) and to the profiles of SCC61 samples ($n = 6$) by rank correlation analysis on the median centered data. The difference between the mean rank correlation to Nu61 and to SCC61 was used as a test statistic. A null distribution of the test statistic for each patient sample was generated by 1,000 random permutations of the Nu61 and SCC61 class labels, and the distribution-free P value for each sample was calculated by comparison to the null distribution. The median P value for each tumor type was less than 0.0001. To compare results from the rank correlation analysis to the k-means support clustering method, samples were assigned as IRDS(+) using the rank analysis if the mean rank correlation was closer to Nu61 than to SCC61 and IRDS(−) if it was closer to SCC61. Although results from both approaches gave highly similar class assignment results, we chose the k-means support clustering method for the breast cancer data. Our preference for this method is because we did not want to impose potential idiosyncrasies of the cell line data on the class assignment process and because of our previous success with the method.

Determination of IRDS Status in Breast Cancer. The unsupervised clustering method described earlier was used to define IRDS status for the 78 patients with breast cancer (NKI78) shown in Fig. 3*A* and Fig. S1. However, this method is not well suited for classification of new samples because it would require a re-analysis of all samples. Therefore, to classify the new samples from the NKI295 data set, we wished to develop an IRDS classifier using supervised class prediction methods trained on the NKI78 data set with the k-means-derived IRDS class assignments. We chose to develop two different classifiers.

Support Vector Machine Classifier. In the first approach, only the 49 IRDS genes were used to develop a support vector machine (SVM) classifier. For this, BRB ArrayTools 3.4.1 and 3.5.0 Beta_1 developed by Richard Simon and Amy Peng Lam (http://linus.nci.nih.gov/BRB-ArrayTools.html) was used with a linear kernel and default tuning parameters and misclassification weights. Our previous work that developed a cell line signature into a clinical classifier using similar approaches as described here suggests that a SVM worked marginally better than other methods such as nearest centroid or k-nearest neighbor (8); however, there is no strong rationale for this, and this decision was primarily arbitrary. Leave-one-out cross-validation resulted in 96% prediction accuracy. Of the 295 patients, 61 are from the 78-patient data set previously described (1). Therefore, 235 unique patients were classified using the SVM classifier and the original IRDS status from the k-means clustering was used for 60 overlapping patients. (Sample 54 from the NKI78 data set had a large proportion of missing values as previously described [8] and was not included in the k-means clustering; hence there were 60 instead of 61 overlapping samples.) It is important to note that, because clinical data were not used to model IRDS class membership, this procedure does not bias analysis of clinical outcome using IRDS status as a factor.

Top Scoring Pairs Classifier. Statistical challenges, such as normalization issues and parameter tuning used in complex classification methods, are potential obstacles to the routine use of gene expression signatures in clinical management. We were motivated to implement a classifier that could facilitate cross-platform use both for validation studies and for clinical application. The TSP classifier (10, 11) has many desirable properties for this purpose. This method uses few genes, has simple decision rules, and is parameter-free because it measures relative gene expression values from pair-wise comparisons.

Using the NKI78 breast cancer patient data set, we used the implementation provided by BRB-ArrayTools 3.5.0 Beta_1 to train a TSP classifier for the k-means-derived IRDS class assignments. Unlike with the SVM classifier, the TSP algorithm selects for gene pairs, necessitating genes besides the 49 IRDS genes. Allowing for a false discovery of only one gene with 99% confidence, there are 162 genes that are differentially expressed between IRDS(+) and IRDS(−) tumors. Although 22 of the 49 IRDS genes are among these 162 genes, even with stringent filtering, the TSP algorithm would likely select gene pairs that did not contain IRDS genes. Indeed, when either no filtering was used or a variance and/or fold change filter was applied, some of the top pairs selected by the TSP method were devoid of an IRDS gene. Therefore, as our goal was not gene discovery and we wanted to avoid potential over-fitting or over-filtering, we restricted the TSP algorithm to using the 49 IRDS genes along with the "intrinsic" breast cancer genes (12). The intrinsic breast cancer genes are 534 genes used to define the molecular subtypes reported by Perou and colleagues (20). From their work, these genes were derived from unsupervised class discovery and showed little variation within the same tumor but high variation between different tumors. The intrinsic genes have been shown to discriminate the different subtypes across different microarray studies and platforms. There were 635 intrinsic breast cancer genes on the NKI78 Agilent microarray platform (duplicate probes were not removed). This was combined with the 49 IRDS genes and probes with missing values in more than one sample were excluded, leaving 648 genes. Thus, we rationalized that the intrinsic genes would be a small set of independent breast cancer genes previously tested across different studies/platforms that could be combined with the IRDS for gene pair selection by the TSP method.

Using the NKI78 as a training set, we selected the number of gene pairs to use by evaluating prediction accuracy using 10-fold cross validation and evaluating an odd number of gene pairs from one to 19. A plateau in prediction accuracy of ≈97% was observed at seven gene pairs; therefore, seven gene pairs were selected. Each gene pair contained an IRDS gene, with the seventh gene pair containing two IRDS genes. For each gene pair, the probability that the IRDS gene has an expression value greater than the non-IRDS gene is greater for IRDS(+) samples. With gene pair seven, in which both genes were from the IRDS, the probability that IFIT3 levels are greater than ZNF273 is greater in IRDS(+) samples, which is consistent with the cell line data showing that ZNF273 is one of the minority of down-regulated genes in the IRDS. A majority vote method was used to train the TSP classifier, meaning that if four of seven IRDS genes scored positive (i.e., had an expression value greater than the non-IRDS gene in each pair), the sample would be classified as IRDS(+). To assess the stability of these seven gene pairs, we added Gaussian noise based on the calculated variance of the training set and ran the TSP algorithm 100 times to select for seven gene pairs. The data were perturbed by sampling from a normal distribution with mean zero and variance equal to the 25th percentile of the calculated variance from the entire data set, and adding this noise to a random sample of 10% of the training set. We chose the 25th percentile rather than the median because a significant proportion of genes were differentially expressed. After 100 runs using the perturbed data sets, the proportion of times each or both of the genes from the original

seven gene pairs were selected was calculated. These results demonstrated that the IRDS genes are reasonably stable and do not significantly fall off until seven gene pairs. The non-IRDS genes from each pair are less stable after the third pair as non-IRDS genes from other pairs can substitute.

Comparison of Different Classifiers. As expected, classification of the NKI295 data set based on the binary TSP classifier was very similar to the SVM results (Cramer V statistic = 0.833). Using an RSF analysis (discussed later), both binary classifiers predicted for metastasis risk after adjusting for other covariates and interactions, and both improved outcome prediction as part of a full model for patients who received adjuvant chemotherapy. In patients who did not receive adjuvant chemotherapy, both binary classifiers had little influence on prediction or accuracy. Thus, although both methods gave highly similar results, we focused our efforts on the TSP because of its clinically desirable properties of using few genes, having simple decision rules, and being parameter-free.

The TSP IRDS Score. Because no clinical data were used in training of either the binary TSP or SVM classifiers, the ability to scale a classifier to clinical outcome could be useful to meet different clinical goals. For example, for a therapy-predictive marker, one could either identify patients most likely to be sensitive to standard adjuvant therapy or patients who are most likely resistant. The seven gene pairs used in the TSP IRDS classifier naturally allows scaling for this purpose. Therefore, rather than use majority vote for binary classification of IRDS status, we used the number of gene pairs that scored positive (TSP IRDS) in the survival and RSF analysis described in the next section.

Survival and RSF Analysis. To preliminarily test the hypothesis that the TSP IRDS is a therapy-predictive marker rather than a prognostic marker, a Cox proportional-hazards regression model was used. In addition to standard clinicopathological factors, the use of adjuvant chemotherapy and interaction terms was modeled to test for an interaction between the TSP IRDS and the use of adjuvant chemotherapy. The proportional-hazards assumption of a Cox regression was tested to ensure that the time-dependent coefficient $\beta(t)$ has a slope of 0 (global $P > 0.05$). Survival analysis was done using the Kaplan-Meier method and the log-rank test. All analysis was done using the survival package 2.26 for R (by T. Therneau and ported by T. Lumley).

RSF Analysis. Recently, there have been several reports on the appropriateness of using a Cox proportional-hazards regression model and reporting hazard ratios with P values to evaluate new tumor markers (13–15). A hazard ratio with a significant P value implies a relationship between the new marker and clinical outcome when other markers remain constant, but it does not directly address the issue of prediction, which is the primary objective. Thus, it has been proposed that models with reasonable assumptions be evaluated with standard markers with and without the new marker of interest. Then, a concordance index or an error rate is used to evaluate prediction accuracy to determine how much the new marker of interest influences the error rate. To apply this paradigm, we used an RSF method to evaluate if the TSP IRDS improves upon outcome prediction. An RSF involves constructing survival trees from bootstrap samples using randomly selected covariates for tree splitting to deliver an ensemble cumulative hazard estimate for survival. It is virtually free of model assumptions and adjusts for each covariate and potential interactions. The RSF also allows the contribution of each covariate to the overall prediction accuracy to be evaluated using an importance score, which measures the change in prediction accuracy of the overall model when the factor is not considered (this is done by permutation). For analysis using RSF, we used randomSurvivalForest package 2.0 (by Hemant Ishwaran and Udaya B. Kogalur) for R. In general, 2,500 to 5,000 survival trees were evaluated. Each of the four splitting rules was tested and the "logrankscore" was used based on consistently better performance compared with the other methods. The default number of predictors was randomly sampled at each split. To estimate prediction accuracy, an out-of-bag error rate (1 − Harrell concordance index) was averaged over 50 to 100 Monte Carlo runs. Importance scores were similarly determined. Models considered included the TSP IRDS with standard clinicopathological markers (i.e., age, tumor size, number of positive lymph nodes, grade, estrogen receptor status), Adjuvant! Online 10-year breast cancer mortality risk, or risk groups determined by St. Gallen classification or genomic classifiers (e.g., NKI 70, wound, molecular subtypes). Partial plots were used to evaluate the influence of each covariate on the expected relative frequency of metastasis.

Determination of TSP IRDS Cut-Off and Survival Analysis. An important clinical goal for a therapy-predictive marker is to identify patients whose disease is likely sensitive to standard adjuvant treatment to avoid over-treatment but also to minimize denying patients with resistant disease more aggressive treatment. From the RSF analysis of the TSP IRDS combined with standard clinicopathological factors or risk groups, a distinct separation in predicted metastasis risk with increasing TSP IRDS is observed between scores less than two and scores of two or greater (Fig. 4 *B* and *D*). When the estimated standard errors are considered, overlap is observed between scores of one and zero, and an examination of the prediction error rate reveals that a score of zero results in higher error than a score of one. Based on these data, a TSP IRDS of less than two was used as a conservative cut-off for Kaplan-Meier survival analysis to identify the patients whose disease was most likely to be sensitive to adjuvant therapy.

The same TSP IRDS cut-off was used in multivariable Cox regression models for importance score calculations along with other covariates that were dichotomized, with the exceptions being tumor size and age, which were treated as continuous variables, and number of positive lymph nodes, if such information was available. Importance scores and prediction error rates were determined using out-of-bag samples. Importance scores were calculated by a "noised-up" method whereby coefficients for each variable was set to zero and the difference in prediction error from out-of-bag samples was measured.

Validation of the IRDS. To validate the properties of the IRDS, several independent breast cancer cohorts were assembled (see Tables S5 and S7). Cohort A is comprised of 292 patients from the Radcliffe, University of California San Francisco, and Stockholm data set who all received adjuvant chemotherapy and/or radiation and was used to validate that the IRDS is a therapy predictive marker for DNA damaging agents, i.e., adjuvant chemotherapy and/or radiation. Recurrence-free survival was used as the endpoint for patients treated with adjuvant chemotherapy and/or radiation and defined as either an LRF or a distant failure as the first event. Distant metastasis-free survival as a first event was used as the endpoint for patients treated with adjuvant chemotherapy, and LRF as a first event was used for patients treated with adjuvant radiation therapy. Cohort B was used to test if the IRDS is specifically a therapy-predictive marker for DNA damaging agents and consists of 277 patients from the Loi data set who received only endocrine therapy for adjuvant systemic treatment. Cohort C was used to validate that the IRDS does not act as a prognostic marker. Analysis of prognostic markers is best done in a manner that minimizes potential influence of treatment on the endpoint studied. Therefore, for cohort C, only 286 patients from the Rotterdam data set who did not receive adjuvant systemic treatment were included,

and distant metastasis-free survival was used as the endpoint. For each of these assembled cohorts, all patients from each data set with enough information were used. Patients were excluded only if the analytic method does not support missing values.

For each data set, calculation of TSP IRDS scores was performed as described for the NKI295 data set. Gene expression values were median centered. In cases in which there were genes not expressed on the particular microarray platform, the corresponding gene pair was omitted and the TSP IRDS score was scaled from zero to seven. For the cross-institute validation analysis, each of four data sets was used as a test set for either a RSF or a Cox model trained on the other three. Importance scores for each covariate were determined by the difference between the error rate using the full model and the error rate with the covariate omitted. Other methods that used random daughter assignment for out-of-bag samples dropped down in-bag survival trees later became available when prediction using separate tests sets was used. In practice, both methods generally yield similar results. For Cox regression, we used the noised-up method. At least 100 to 1,000 Monte Carlo runs were used for each test set and the importance score was averaged. When using RSF, missing data were imputed. For Cox regression, missing data were excluded.

To provide average importance score values and estimate their sampling variability, all breast cancer data sets from Table S8 were combined and overlapping patients were excluded. This resulted in 1,573 patients. Both RSF and Cox models were used. However, with RSF missing, data were imputed and analysis was non-stratified. Because all possible variable interactions were considered, treatment effect could be extracted. Importance scores were calculated by random daughter node assignment for out-of-bag samples dropped down in-bag survival trees. In contrast, Cox models were stratified by treatment and no interactions were included. Bootstrap means and SEs for both importance scores and prediction error rates were determined using at least 100 to 1,000 bootstrap samplings with replacement.

1. van de Vijver MJ, *et al.* (2002) A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347:1999–2009.
2. Chang HY, *et al.* (2005) Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci USA* 102:3738–3743.
3. Fan C, *et al.* (2006) Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 355:560–569.
4. McShane LM, *et al.* (2002) Methods for assessing reproducibility of clustering patterns observed in analyses of microarray data. *Bioinformatics* 18:1462–1469.
5. Goldhirsch A, *et al.* (2005) Meeting highlights: international expert consensus on the primary therapy of early breast cancer 2005. *Ann Oncol* 16:1569–1583.
6. Torres-Roca JF, *et al.* (2005) Prediction of radiation sensitivity using a gene expression classifier. *Cancer Res* 65:7169–7176.
7. Khodarev NN, *et al.* (2004) STAT1 is overexpressed in tumors selected for radioresistance and confers protection from radiation in transduced sensitive cells. *Proc Natl Acad Sci USA* 101:1714–1719.
8. Minn AJ, *et al.* (2005) Genes that mediate breast cancer metastasis to lung. *Nature* 436:518–524.
9. Saeed AI, *et al.* (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34:374–378.
10. Tan AC, Naiman DQ, Xu L, Winslow RL, Geman D (2005) Simple decision rules for classifying human cancers from gene expression profiles. *Bioinformatics* 21:3896–3904.
11. Xu L, Tan AC, Naiman DQ, Geman D, Winslow RL (2005) Robust prostate cancer marker genes emerge from direct integration of inter-study microarray data. *Bioinformatics* 21:3905–3911.
12. Sorlie T, *et al.* (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 100:8418–8423.
13. Kattan MW (2004) Evaluating a new marker's predictive contribution. *Clin Cancer Res* 10:822–824.
14. Katz EM, Kattan MW (2005) How to judge a tumor marker. *Nat Clin Pract Oncol* 2:482–483.
15. Simon R (2005) Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol* 23:7332–7341.
16. van 't Veer LJ, *et al.* (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415:530–536.
17. Chin K, *et al.* (2006) Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* 10:529–541.
18. Korkola JE, *et al.* (2007) Identification of a robust gene signature that predicts breast cancer outcome in independent data sets. *BMC Cancer* 7:61.
19. Pawitan Y, *et al.* (2005) Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* 7:R953–R964.
20. Sotiriou C, *et al.* (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci USA* 100:10393–10398.
21. Wang Y, *et al.* (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365:671–679.
22. Ma XJ, *et al.* (2004) A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 5:607–616.
23. Sotiriou C, *et al.* (2006) Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 98:262–272.
24. Loi S, *et al.* (2007) Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* 25:1239–1246.
25. Desmedt C, *et al.* (2007) Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res* 13:3207–3214.
26. Chung CH, *et al.* (2004) Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. *Cancer Cell* 5:489–500.
27. Bild AH, *et al.* (2006) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439:353–357.
28. Stephenson AJ, *et al.* (2005) Integration of gene expression profiling and clinical variables to predict prostate carcinoma recurrence after radical prostatectomy. *Cancer* 104:290–298.
29. Freije WA, *et al.* (2004) Gene expression profiling of gliomas strongly predicts survival. *Cancer Res* 64:6503–6510.
30. Phillips HS, *et al.* (2006) Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 9:157–173.

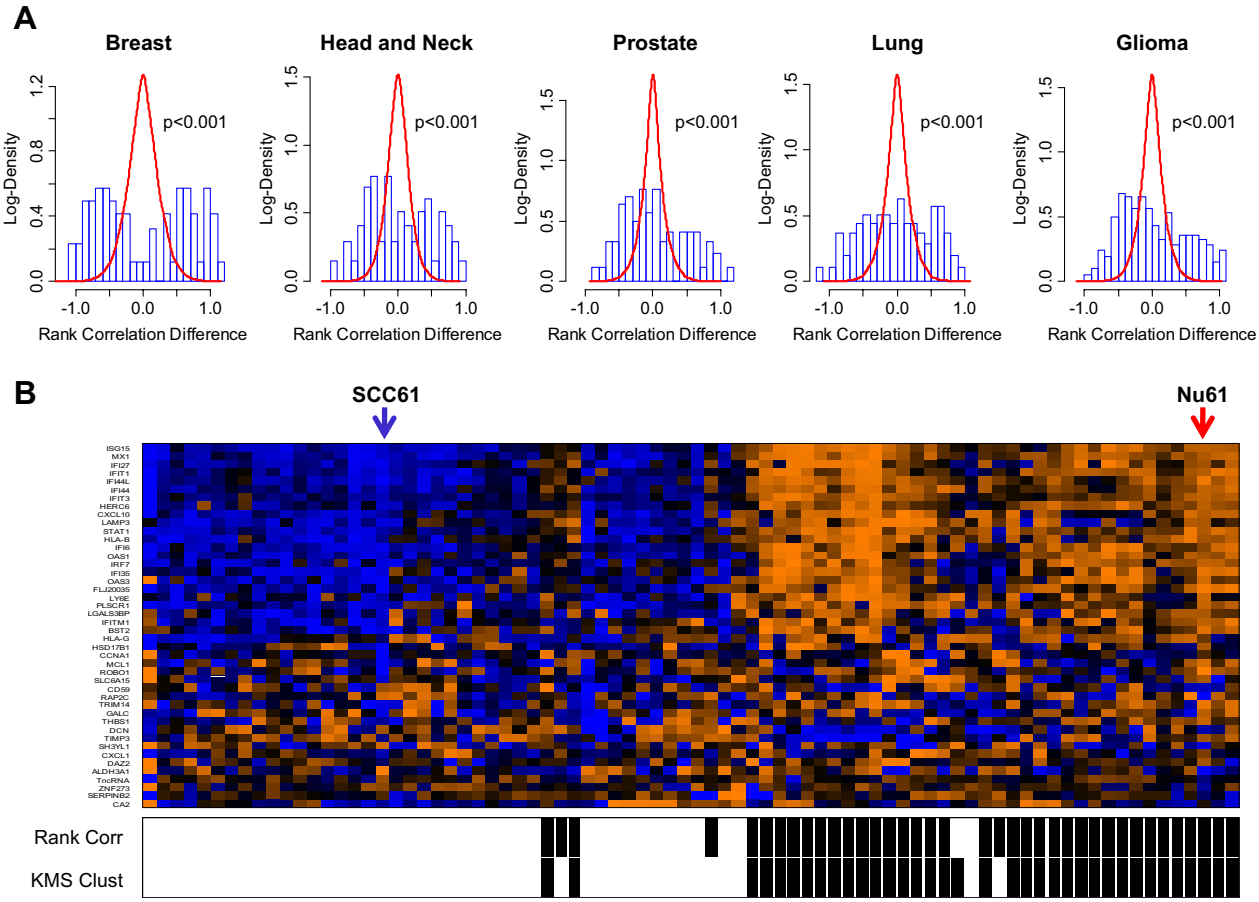

Fig. S1. Rank correlation analysis comparing IRDS gene expression profiles of primary human cancers versus Nu61/SCC61 and class assignment of 78 breast tumors. (*A*) The gene expression profile of the IRDS genes for each patient sample from the indicated primary tumor types was compared with the IRDS gene profiles of Nu61 samples and to the profiles of SCC61 samples by rank correlation analysis. The difference between the mean rank correlation to Nu61 and to SCC61 was defined as the rank correlation difference (blue histogram) and was used as a test statistic. Positive values of the rank correlation difference represent similarity to the Nu61 IRDS gene profiles and negative values represent similarity to SCC61. The Nu61 profile defines the IRDS(+) state and that of SCC61 defines the IRDS(−) state. A null distribution of the test statistic was generated by random permutation of the Nu61 and SCC61 class labels. The distribution-free *P* value for each sample was calculated by comparison to the null distributions. Shown are the null distributions (red) and the median *P* values for the indicated cancer types. (*B*) The IRDS gene expression profiles for 78 patients with breast cancer along with the Nu61 and SCC61 IRDS centroid are shown in the heat map with samples in columns and IRDS genes in rows. Orange is high gene expression and blue is low. The position of the Nu61 and the SCC61 centroids are indicated by arrows. IRDS(+) and IRDS(−) class assignments were made by comparing each patient sample to the Nu61 and SCC61 IRDS gene profiles. Samples were assigned as IRDS(+) if the mean rank correlation was closer to Nu61 than to SCC61 and IRDS(−) if it was closer to SCC61. Alternatively, k-means support clustering with a k value of 2 was performed using the IRDS genes to divide the patient samples into two groups. The centroid of each group was compared with the Nu61 and the SCC61 centroid. The group positively correlating with Nu61 was assigned as IRDS(+) and the group positively correlating with SCC61 was assigned as IRDS(−). Results of each method are shown on the bottom of the heat map, with vertical black bars indicating IRDS(+) status. The correlation between the two methods is 0.90.

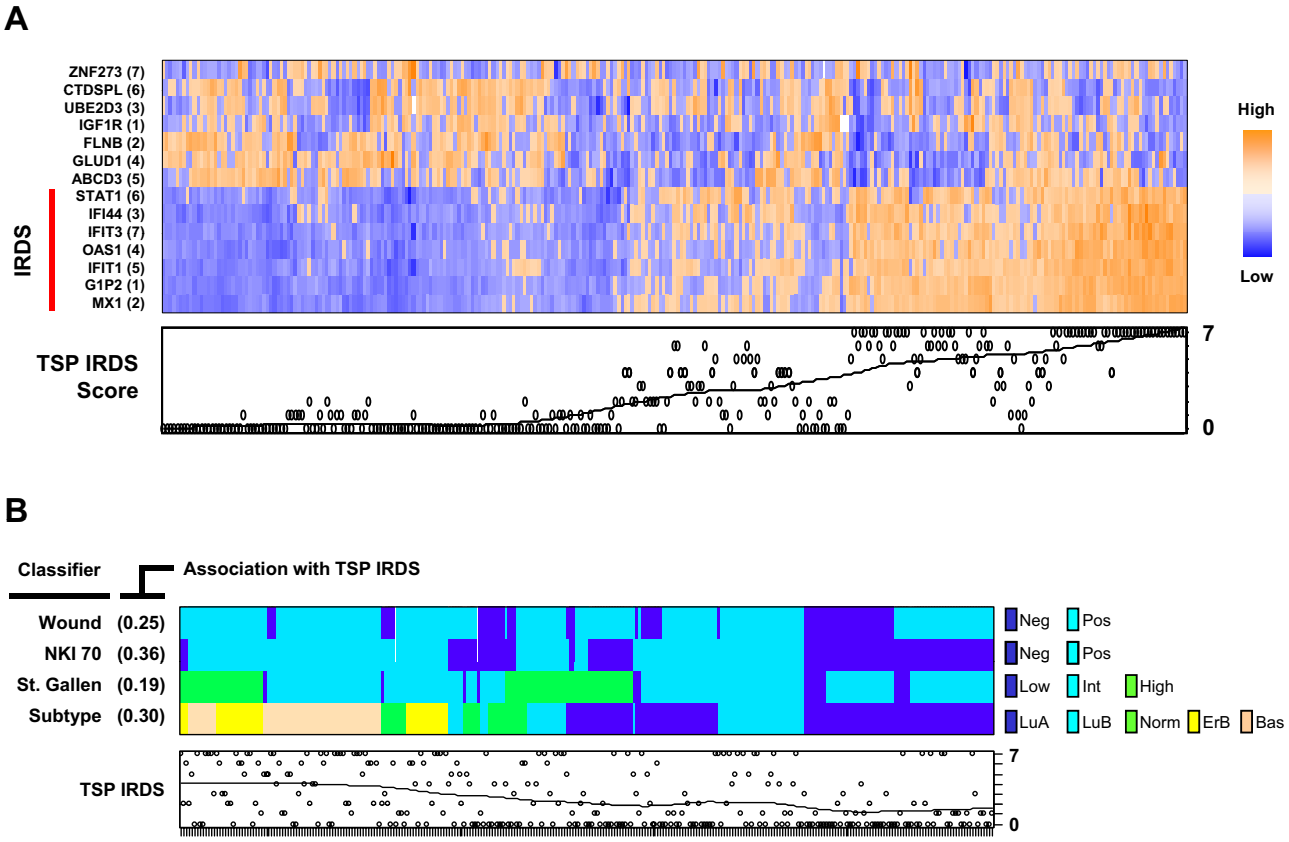

Fig. S2. Expression of the TSP IRDS score among 295 primary breast tumors. (*A*) The expression pattern of the seven gene pairs in the TSP classifier for IRDS status is shown for the NKI295 data set. Each IRDS gene (delineated by red line) and its partner gene used in a pair-wise comparison are numbered in parentheses. Each column is a primary tumor and each row is a gene, with orange representing high expression and blue low expression. The number of gene pairs that score with the TSP method (TSP IRDS) are shown below the heat map. (*B*) Association between patients classified according to the indicated clinical or genomic classifier is shown by hierarchical clustering using the NKI295 breast cancer data set. Each vertical tick represents a patient. The legend for the color codes for each group is shown on the right of the heat map. The TSP IRDS score is displayed below the heat map. The association between classification by the indicated risk groups and the TSP IRDS score is measured by a Cramer V statistic and shown on the left in parentheses (a value of 0 indicates no relationship and values from 0.20 to 0.49 indicate a moderate to substantial relationship). In general, worse prognosis has been reported for patients who are are wound signature (wound)-positive, NKI 70-positive, at high risk per St. Gallen criteria, or belong to the ERBB2/Her2 (ErB) or Basal-like (Bas) molecular subtypes as compared to the luminal A (LuA) subtype. Intermediate-risk subtypes include the luminal B (LuB) and normal-like (Norm) subtypes.

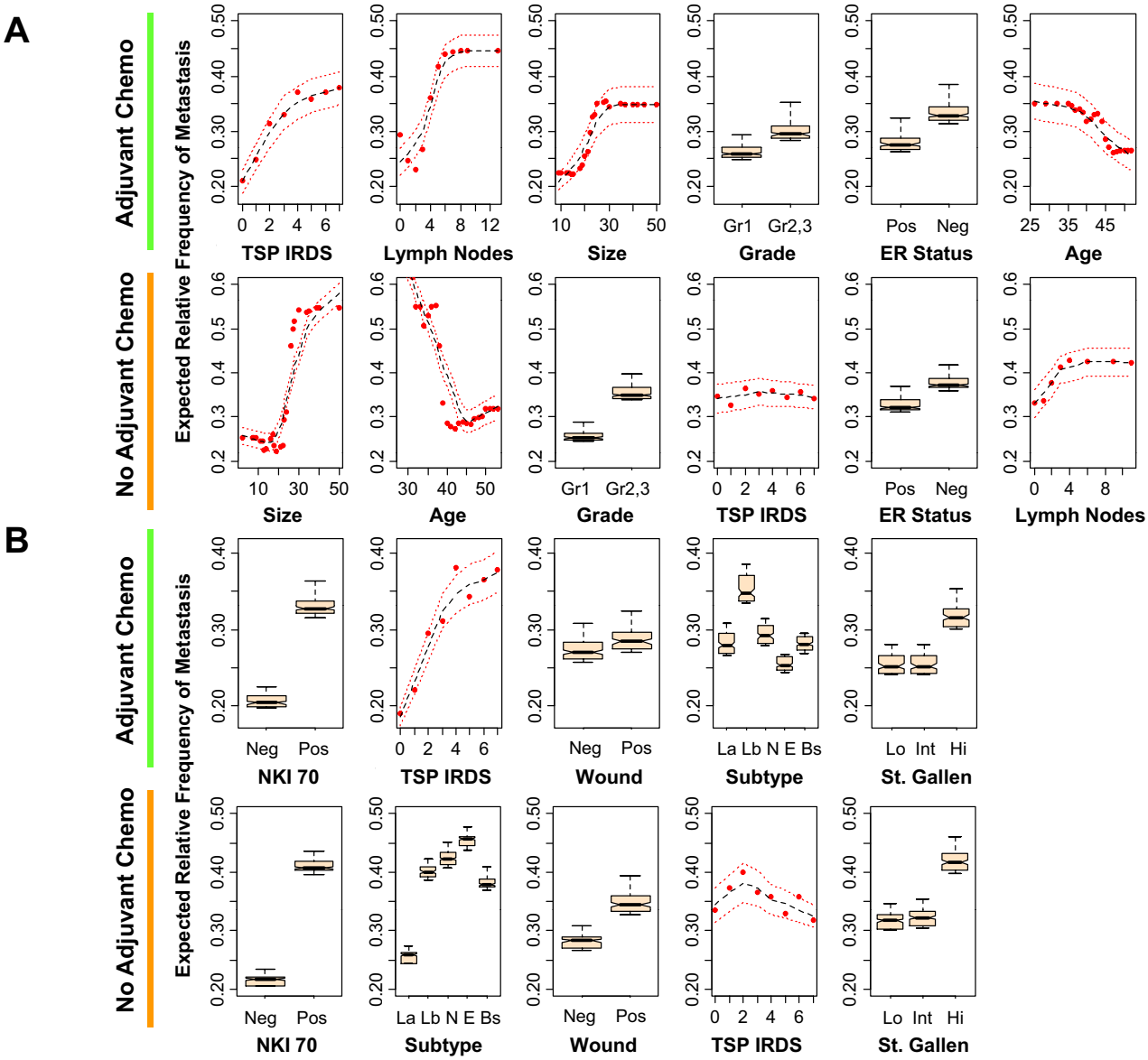

Fig. S3. Partial plots of the influence of the TSP IRDS, standard clinicopathological markers, genomic markers, and risk groups on metastatic risk from multivariable RSF models. The 110 patients treated with adjuvant chemotherapy (grouped beside green bar) and the 185 patients not treated with adjuvant chemotherapy (grouped beside orange bar) from the NKI295 data set were separately analyzed using an RSF analysis as described in Fig. 4. The TSP IRDS was combined with (*A*) standard clinicopathological factors or (*B*) clinical risk groups defined by the 2005 St. Gallen criteria or other gene expression signatures that include the NKI 70, wound, and molecular subtype classification (La, luminal A; Lb, luminal B; N, normal-like; E, ERBB2/Her2; Bs, basal-like). These partial plots show the expected relative frequency of patients developing metastasis as a function of each of the indicated covariates after adjusting for all other covariates in the model. For continuous variables, the estimated risk is shown (red dot) along with a Lowess regression (black dashes) ± 2SE (red dashes). Discrete variables are shown by box-and-whisker plots with non-overlapping notches considered significant. The covariates are ordered from left to right by rank of their importance score. See Fig. 4.

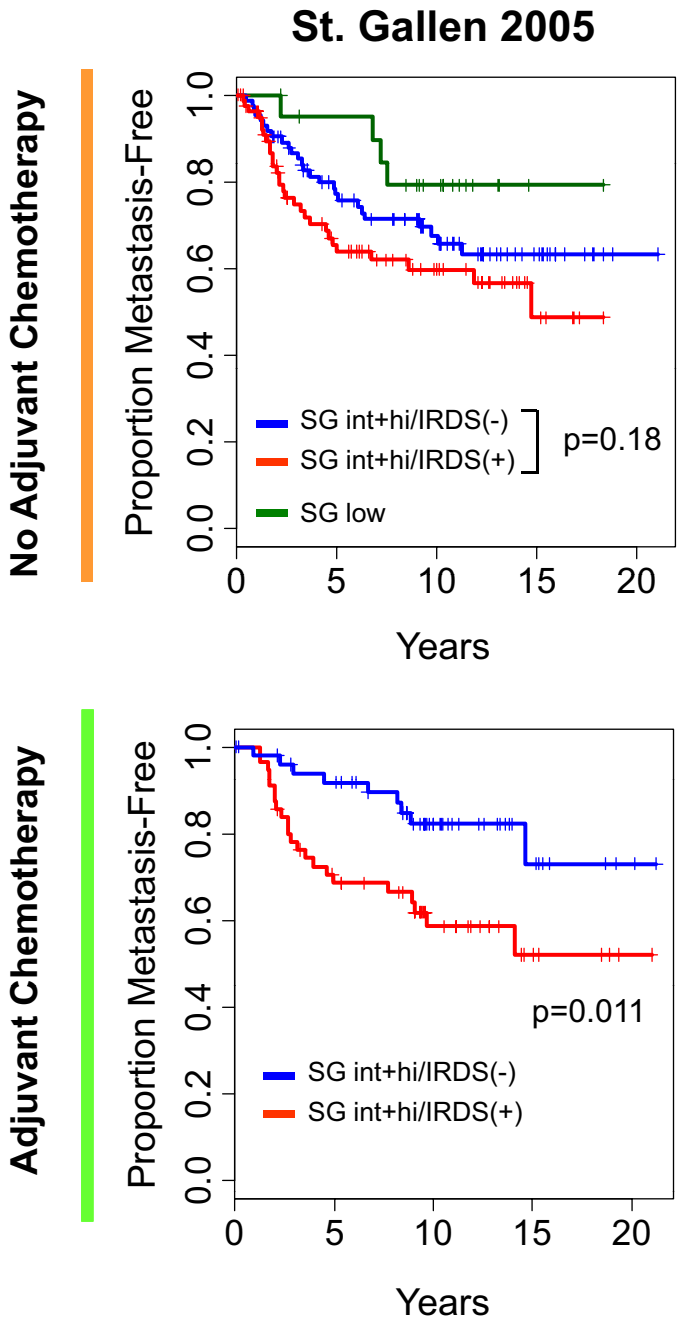

Fig. S4. The IRDS identifies patients at high risk per St. Gallen criteria who are rendered low risk with adjuvant chemotherapy. Each of the 185 patients who did not receive ADCT (grouped beside the orange line) or the 110 patients who received ADCT (grouped beside the green line) were classified using the 2005 St. Gallen criteria as having a good prognosis (SG low) or a poor prognosis (SG int+hi). The patients with a poor prognosis, who are typically considered for ADCT, were further split by IRDS status. A TSP IRDS score of <2 and ≥ 2 were used to define IRDS(−) and IRDS(+), respectively. Shown are the metastasis-free survival curves. The log-rank *P* values shown are for comparison between the groups stratified by IRDS.

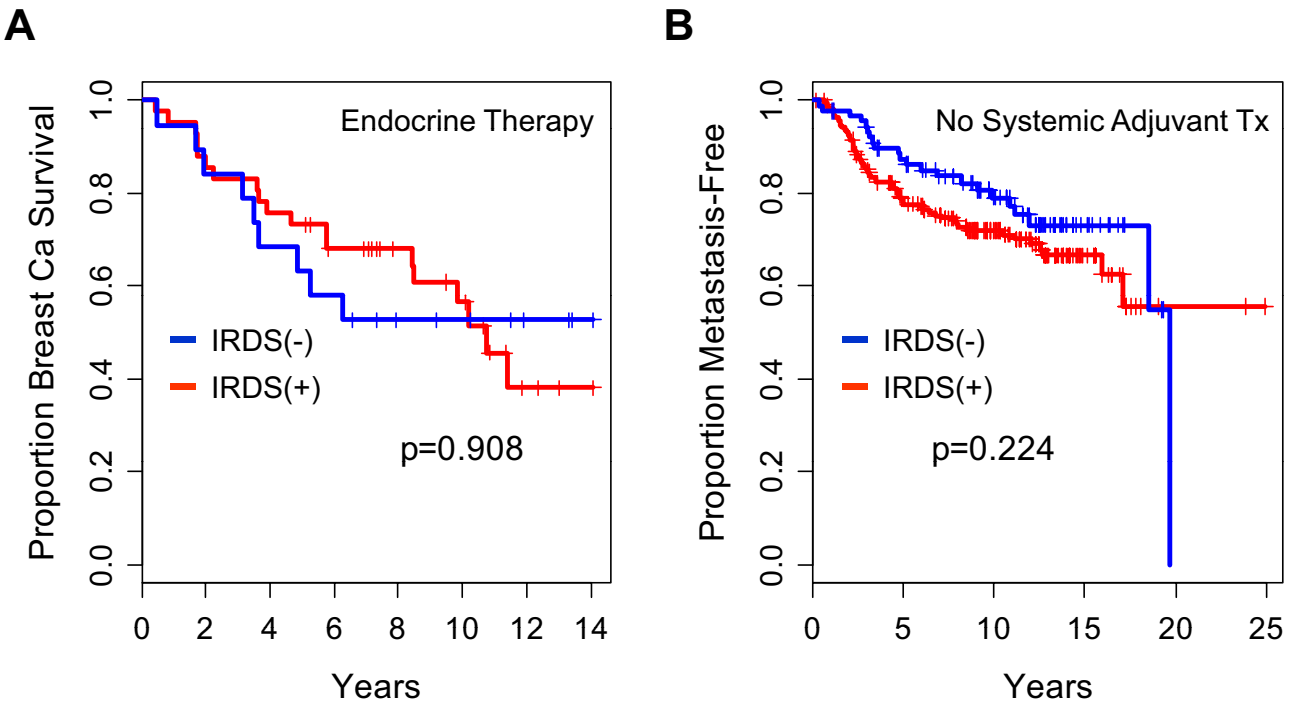

Fig. S5. The IRDS is not a therapy-predictive marker for endocrine therapy or a prognostic marker in the absence of systemic treatment. (*A*) Breast cancer survival for 60 patients (MGH) receiving endocrine therapy and (*B*) metastasis-free survival for 282 patients (TRANSBIG and KJ125, non-overlapping) not receiving systemic adjuvant treatment were analyzed by IRDS status. All *P* values are calculated by the log-rank test.

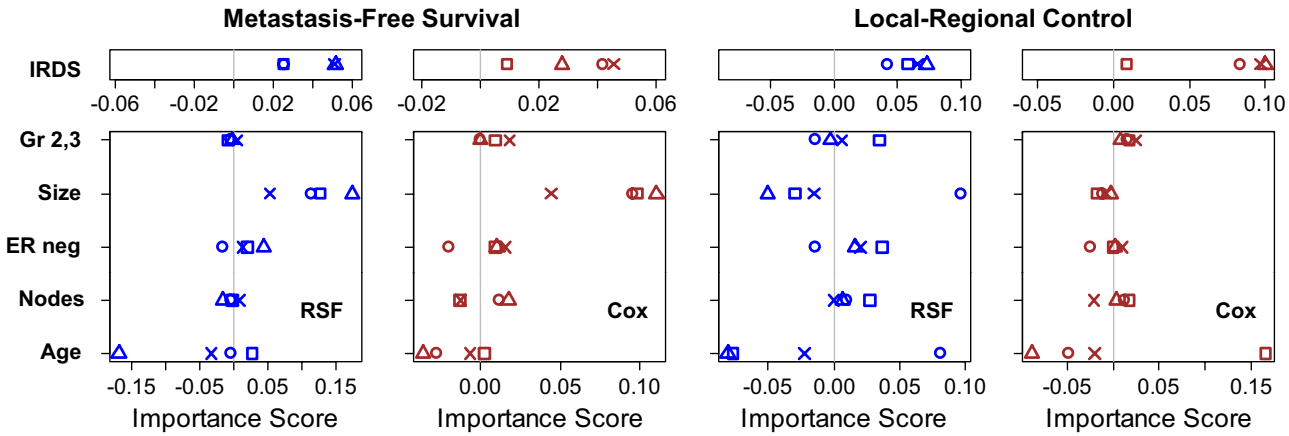

Fig. S6. Cross-institute analysis of the IRDS as a therapy predictive marker for adjuvant chemotherapy and radiation. Patients who received either adjuvant chemotherapy or radiation from the NKI295, University of California San Francisco, Radcliffe, and Stockholm data sets were used in an external validation study for the TSP IRDS. For this, metastasis-free survival (MFS), for patients receiving chemotherapy, or local-regional control (LRC), for patients receiving radiation therapy, were predicted from each data set using either an RSF (blue) or Cox (brown) model trained on the other three data sets. Shown are importance scores of the indicated covariates for each data set when used as a test set (Stockholm, square; University of California San Francisco, circle; Radcliffe, triangle; NKI295, cross).

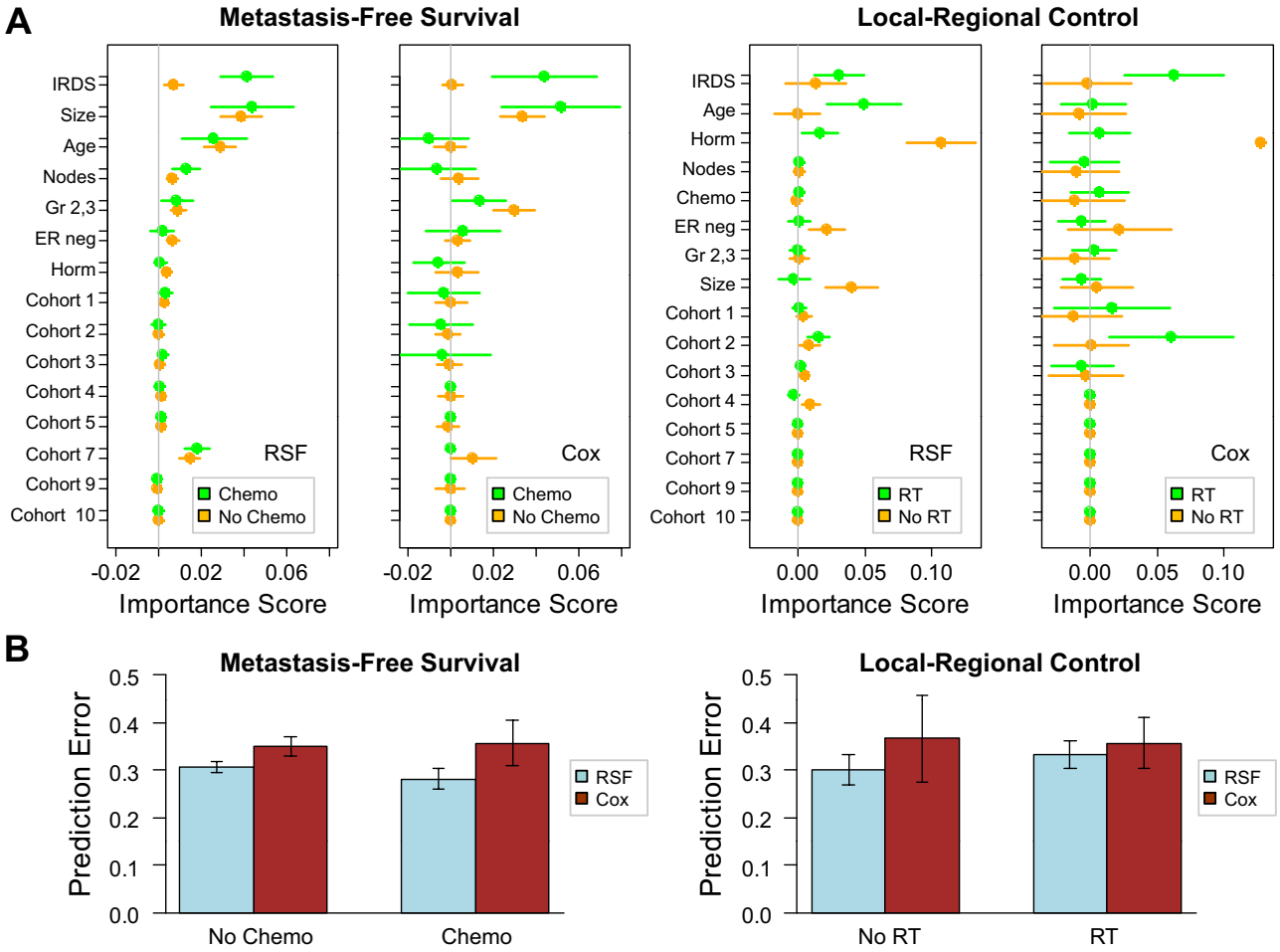

Fig. S7. Metaanalysis of the IRDS as a therapy-predictive marker for adjuvant chemotherapy and/or radiation. (*A*) All data sets used in validation were combined with the NKI295, resulting in a merged set of 1,573 patients. This merged set was used in a metaanalysis to estimate means and SEs for the importance scores of the indicated covariates for metastasis-free survival (MFS) and local-regional control (LRC) based on bootstrap resampling. For RSF, missing values were imputed and a non-stratified model incorporating non-linear effects and multi-way interactions was used to extract treatment effect. In this way, cohort effects could also be determined. The cohorts contributed little or no effect on prediction, indicating that an adequate level of homogeneity across institutes, and the analysis accounted for differences in these cohorts. For Cox regression, analysis was stratified by treatment and variable interactions were not modeled. (*B*) The bootstrap prediction errors (mean ± SD) for the RSF models and the Cox models for MFS and LRC in the absence of treatment and in the presence of treatment are shown.

Table S1. Multivariable Cox proportional-hazards model of distant metastasis for NKI295 breast cancer patients

Variable	HR	95% CI	*P* value
Chemotherapy	0.35	0.17–0.72	0.004
TSP IRDS, per point, 0–7	0.97	0.49–1.91	0.930
ER negative	1.82	0.83–3.98	0.140
Grade 2/3	3.98	1.53–10.40	0.005
Age, per y	0.94	0.89–1.00	0.045
Nodes positive, per node	1.20	1.08–1.34	0.001
Tumor size, per mm	1.04	1.01–1.07	0.021
TSP IRDS interaction with			
Chemotherapy	1.20	1.00–1.43	0.050
ER negative	0.94	0.79–1.11	0.460
Grade 2/3	0.81	0.63–1.03	0.090
Age	1.01	0.99–1.02	0.330
Nodes positive	0.97	0.94–1.00	0.061
Tumor size	1.00	0.99–1.01	0.810

HR, hazard ratio.

Table S2. Characteristics of the NKI295 breast cancer patients according to TSP IRDS score

	TSP IRDS score								
Variable	0	1	2	3	4	5	6	7	Total	*P* value*
Grade										0.001
1	43	8	8	4	3	4	3	2	75	
2/3	73	26	17	10	14	13	17	50	220	
ER status										<0.001
Positive	104	29	15	11	13	10	11	33	226	
Negative	12	5	10	3	4	7	9	19	69	
Lymph nodes										0.40
Negative	62	20	8	6	11	10	10	24	151	
Positive	54	14	17	8	6	7	10	28	144	
Tumor size										0.068
T1	68	19	12	7	11	12	7	19	155	
T2+	48	15	13	7	6	5	13	33	140	
Age, y										0.89
<30	22	6	6	2	5	6	6	10	63	
30–39	74	23	13	10	9	10	10	34	183	
40–50+	20	5	6	2	3	1	4	8	49	

*χ^2 test.

Table S3. Characteristics of the NKI295 breast cancer patients stratified by adjuvant chemotherapy and radiation

Variable	No chemotherapy ($n = 185$)	Adjuvant chemotherapy ($n = 110$)	*P* value	Adjuvant RT ($n = 243$)
Mean age, y	44.1	43.7	0.53	44.0
Mean size, cm	2.19	2.37	0.09	2.17
Grade 2/3, %	75.1	73.6	0.88	72.0
Node positive, %	21.6	94.5	<0.001	57.6
ER negative, %	25.9	19.1	0.23	22.2
Chemotherapy, %	0	100		44.0
Hormones, %	10.8	18.1	0.11	14.4

For categorical variables, *P* values were calculated by a χ^2 test, and for continuous variables *P* values were calculated by either a Student *t* test or by a Wilcoxon rank-sum test for skewed data.

Table S4. Multivariable Cox proportional-hazards model of LRF for NKI295 patients treated with adjuvant radiation (n = 243)

Variable	HR (95% CI)	P value
IRDS, TSP IRDS <2 vs. ≥2	0.37 (0.17–0.80)	0.012
Grade 2/3	1.28 (0.52–3.12)	0.59
Age, per y	0.90 (0.85–0.96)	0.001
Nodes positive, per node	1.04 (0.86–1.26)	0.67
Tumor size, per mm	0.98 (0.94–1.03)	0.44
ER negative	1.23 (0.55–2.75)	0.62
Mastectomy	0.29 (0.09–0.91)	0.034
Adjuvant chemotherapy	1.08 (0.46–2.53)	0.87

HR, hazard ratio.

Table S5. Patient characteristics for cohorts used in validation studies

Variable	Cohort A (n = 292)			Cohort B (n = 277)	Cohort C (n = 286)
	All patients	Adjuvant chemotherapy (n = 140)	Adjuvant RT (n = 242)		
Mean age, y	54.4	49.2	55.7	64.2	53.9
Mean size, cm	2.58	2.80	2.50	2.55	2.32
Grade 2/3, %	86.8	92.8	86.1	78.1	96.4
Node positive, %	58.8	77.0	57.4	55.2	0
ER negative, %	31.0	39.3	28.8	1.9	26.9
Chemotherapy, %	47.9	100	37.2	0	0
Hormones, %	67.4	53.2	71.4	100	0
Radiation, %	82.9	64.3	100	NA	NA

Cohort A includes the University of California San Francisco, Radcliffe, and Stockholm groups and was used to validate the IRDS as a therapy predictive marker for adjuvant chemotherapy and/or radiation. Cohort B includes the Loi group and was used to test specificity as a therapy predictive marker for DNA damaging agents. Cohort C includes the Rotterdam group and was used validate that the IRDS is not a prognostic marker. NA = not available.

Table S6. Univariate Cox proportional-hazards model of recurrence-free survival for patients receiving adjuvant chemotherapy and/or RT in the NKI295 data set and three validation data sets.

Data set	HR (per point, 0–7)	95% CI	*P* value
NKI, $n = 246$	1.11	1.04–1.19	0.002
Radcliffe, $n = 84$	1.24	1.07–1.43	0.003
UCSF, $n = 119$	1.16	0.994–1.34	0.060
Stockholm, $n = 89$	1.18	1.00–1.40	0.047

HR, hazard ratio; UCSF, University of California San Francisco.

Table S7. Patient characteristics for each breast cancer data set used in cross-validation studies

Variable	NKI	Radcliffe	UCSF	Stockholm
Chemotherapy				
No. of patients	110	32	78	30
Mean age, y	43.7	49.3	49.9	47.3
Mean tumor size, cm	2.37	3.07	2.82	2.45
Grade 2/3, %	73.6	96.9	93.5	86.7
Node positive, %	94.5	84.4	68.8	90.0
ER negative, %	19.1	59.4	41.0	13.3
Hormones, %	18.2	68.8	44.2	60.0
Radiation, %	97.3	87.5	56.4	60.0
CMF, %	98.2	90.6		80.0
Anthracycline-containing, %	1.8	9.4	″primarily″	20.0
Radiation				
No. of patients	243	80	85	77
Mean age, y	44	56	55.1	55.9
Mean tumor size, cm	2.17	2.75	2.57	2.17
Grade 2/3, %	72.0	85.0	91.7	80.8
Node positive, %	57.6	57.5	55.3	59.7
ER negative, %	22.2	35.0	32.5	18.2
Hormones, %	14.4	81.3	58.3	75.3
Chemotherapy, %	44.0	35.0	51.8	23.4

CMF, cyclophosphamide/methotrexate/5-fluorouracil; UCSF, University of California San Francisco.

Table S8. Microarray data sets used in this study by cancer type and alias for the data set referred to in the study

Cancer type	Alias	No. of samples	Reference(s)
Breast	NKI78	78	16
Breast	NKI295	295	1
Breast	UCSF	175	17,18
Breast	Stockholm	159	19
Breast	Radcliffe	99	20
Breast	Rotterdam	286	21
Breast	MGH	60	22
Breast	KJ125	125	23
Breast	Loi	277	24
Breast	TRANSBIG	198	25
Head and neck	–	60	26
Lung	–	86	27
Prostate	–	78	28
High-grade glioma	–	185	29,30

Table S9. Probe identifiers for 49 IRDS genes by gene symbol

Gene symbol	Affymetrix probe set ID	NKI unique ID	Chung unique ID
ALDH3A1	205623_at	NM_000691	10605
BST2	201641_at	NM_004335	NA
CA2	209301_at	NM_000067	7901
CCNA1	205899_at	NM_003914	5249
CD59	212463_at	NM_000611	14065
CXCL1	204470_at	NM_001511	7799
CXCL10	204533_at	NM_001565	5020
DAZ1	216922_x_at	AF248480	9882
DCN	201893_x_at	NM_001920	11778
FLJ20035	218986_s_at	NM_017631	NA
FLJ38348	213294_at	NA	NA
G1P2	205483_s_at	NM_005101	5740
G1P3	204415_at	NM_002038	16530
GALC	204417_at	NM_000153	NA
HERC6	219352_at	NM_017912	15099
HLA-B	208729_x_at	NM_005514	9374
HLA-G	211529_x_at	NM_002127	6816
HSD17B1	205829_at	NM_000413	7619
IFI27	202411_at	NM_005532	9130
IFI35	209417_s_at	U72882	8878
IFI44	214453_s_at	NM_006417	5973
IFI44L	204439_at	NM_006820	15590
IFIT1	203153_at	NM_001548	NA
IFIT3	204747_at	NM_001549	10164
IFITM1	201601_x_at	NM_003641	15086
IGF2	202409_at	NA	NA
IRF7	208436_s_at	NM_004029	10241
LAMP3	205569_at	NM_014398	NA
LGALS3BP	200923_at	NM_005567	12963
LY6E	202145_at	NM_002346	7243
MCL1	200798_x_at	L08246	4944
MX1	202086_at	NM_002462	8263
MX2	204994_at	NA	7039
OAS1	205552_s_at	NM_016816	17219
OAS3	218400_at	NM_006187	NA
OASL	205660_at	NA	NA
PLSCR1	202446_s_at	AB006746	6639
RAP2C	218669_at	NM_021183	314
ROBO1	213194_at	NM_002941	5321
SERPINB2	204614_at	NM_002575	16572
SH3YL1	204019_s_at	NM_015677	5324
SLC6A15	206376_at	NM_018057	3904
STAT1	209969_s_at	NM_007315	NA
THBS1	201110_s_at	NM_003246	6554
TIMP3	201150_s_at	NM_000362	17867
TncRNA	214657_s_at	U60873	3830
TRIM14	203148_s_at	NM_014788	9634
USP18	219211_at	NA	7801
ZNF273	215239_x_at	X78932	11704

The Affymetrix U133A platform was used in the prostate, lung, and glioma data sets and the probe set identifiers are shown. Also shown are the unique identifiers for the microarray platforms used for the NKI78 breast cancer data set and the head and neck data set of Chung *et al.* (ref. 26 in *SI Text*). NA, no corresponding probes.

Table S10. shRNAmir sequences for IRDS gene targeting. The targeting sequence for each of the indicated genes is shown

Gene	Targeting Sequence
STAT1	TGCTGTTGACAGTGAGCGCGGCCCTAAAGGAACTGGATATTAGTGAAGCCACAGATGTA
	ATATCCAGTTCCTTTAGGGCCATGCCTACTGCCTCGGA
IFIT1	TGCTGTTGACAGTGAGCGCGGCATCATTAACAAGGGATAATAGTGAAGCCACAGATGTA
	TTATCCCTTGTTAATGATGCCTTGCCTACTGCCTCGGA
ISG15	TGCTGTTGACAGTGAGCGCGAGCATCCTGGTGAGGAATAATAGTGAAGCCACAGATGTA
	TTATTCCTCACCAGGATGCTCATGCCTACTGCCTCGGA
NS	TGCTGTTGACAGTGAGCGATCTCGCTTGGGCGAGAGTAAGTAGTGAAGCCACAGATGTA
	CTTACTCTCGCCCAAGCGAGAGTGCCTACTGCCTCGGA

NS, non-silencing control sequence.